DOES GOD PLAY DICE? ROGER PENROSE, QUANTUM CONSCIOUSNESS, AND THE DEBATE OVER THE LIMITS OF SCIENCE

A talk delivered at the conference "Science and Religion: The Religious Beliefs and Practices of Scientists—20th Century," Institut für Wissenschaftsgeschichte, Georg-August-Universität, Göttingen, 28 May 2002

Gabriel Finkelstein

University of Colorado at Denver

Introduction

I have been told that David Hilbert, one of the greatest mathematicians of the twentieth century, is buried right here in Göttingen. Apparently, on his tombstone there is an inscription drawn from the last sentence of a speech that he broadcast from his hometown of Königsberg on 8 September 1930.[1] It reads: *Wir müssen wissen—wir werden wissen!*[2] This exhortation recalls the final word of a similar lecture delivered forty-eight years earlier by the physiologist and philosopher Emil du Bois-Reymond. Speaking in Leipzig to the 45th plenary session of the Society of German Scientists and Physicians, du

[1] http://topo.math.u-psud.fr/~lcs/Hilbert/HlbrtKD.htm
[2] "We must know—we will know!"

Bois-Reymond broached the topic "Of the Limits of Natural Knowledge" and declared, at least with respect to the riddle of consciousness, *Ignorabimus*—we will never know.[3]

I bring up these famous pronouncements because they set the stage for my remarks today. On the one hand, we have du Bois-Reymond's doubt that science will be able to address the most important questions of existence. On the other hand, we have Hilbert's belief that the methods of science are the only means of ever obtaining those answers. The subject of my present talk, the mathematician and physicist Roger Penrose, has sided with Hilbert on this issue. I would like to suggest that Penrose's current position on the nature of consciousness makes most sense in the context of this long-standing debate on the limits of science, and what is more, the even longer-standing debate between science and religion.

But first let us first turn to Penrose's argument.

Gödel, Turing, and the Non-Computability of Mathematical Reasoning

In 1905, following the lead of du Bois-Reymond, Hilbert identified twenty-three unsolved problems in mathematics.[4] By 1928 he expanded the tenth in this list to what he termed the *Entscheidungsproblem*: Is there a general formal procedure, or algorithm, by which mathematicians can determine if their assertions are true? Answers to Hilbert's question soon arrived. In 1931 Kurt Gödel proved that all consistent axiomatic systems must be incomplete. Five years later Alan Turing demonstrated that some mathematical questions can never be decided by calculation. Hilbert's hopes of finding a formal foundation to mathematics seem to have foundered.

---

[3] Emi du Bois-Reymond, "Über die Grenzen des Naturerkennens" (14 August 1872), in *Reden*, 2nd edn., 2 vols. (Leipzig: Verlag von Veit & Comp., 1912), *1*: 441-473.
[4] http://aleph0.clarku.edu/~djoyce/hilbert/problems.html

Jump to 1961. The formal device Turing imagined as trying to solve Hilbert's *Entscheidungsproblem* now exists: it's called a computer. Reflecting on the difference between men and machines, the Oxford philosopher John Lucas borrows Gödel's proof to argue that consciousness can never be modeled by algorithms. He meets with several potent criticisms. There the matter lies more or less dormant until 1989, when Roger Penrose revisits Lucas's argument in *The Emperor's New Mind*, a popular survey of his contentions with artificial intelligence, quantum mechanics, and theories of consciousness.[5] Five years later Penrose publishes a sequel, *Shadows of the Mind*, which consists largely of responses to criticism of his argument.[6] His third, highly derivative summary appears in 1997.[7] Today Penrose continues to debate his position in a variety of scholarly and popular journals.[8]

Let me now attempt an epitome of Penrose's thesis. Although I have an undergraduate degree in physics, I found much of Penrose's discussion tough going. He has a broad and deep knowledge of science and mathematics, and even his popular writings treat technical issues of computability theory, Mandelbrot sets, quasi-periodic tilings, special and general relativity, quantum mechanics, cosmology, and neuroscience. I am expert in none of these areas; in fact, it took me more than three weeks of immersion in his work to make sense of his position. I think I understand it pretty well now, but if there are any physicists or logicians in the audience horrified at the hash I'm about to make of my subject, I beg your forbearance. On the other hand, if my summary sounds

---

[5] *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics* (Oxford: Oxford University, 1989).
[6] *Shadows of the Mind: A Search for the Missing Science of Consciousness* (Oxford: Oxford University, 1994).
[7] Roger Penrose, Abner Shimony, Nancy Cartwright, and Stephen Hawking, *The Large, the Small, and the Human Mind* (Cambridge: Cambridge University, 1997).
[8] Roger Penrose, "Beyond the Doubting of a Shadow: A Reply to Commentaries on *Shadows of the Mind*," *Psyche* 2.23 (January 1996); Roger Penrose and Stuart Hameroff, "What 'Gaps'? Reply to Grush and Churchland," *Journal of Consciousness Studies* 2 (1995): 99-112.

like utter gibberish to you, I recommend that you just nod and smile through this part of the talk. I promise to return to history in the conclusion, and nodding and smiling has always worked for me.

Here, in extremely abbreviated form, is why Penrose thinks we need a new theory of physics:

1. Consciousness cannot be modeled by any formal, consistent system. It therefore must be nonalgorithmic.
2. Consciousness is a reality that deserves scientific explanation.
3. No current theory of physics, not even quantum mechanics, is nonalgorithmic.
4. Therefore, a new physical theory of quantum gravity is needed to explain the action of the mind.

Let me now expand on each of these points. The first one, which treats the nonalgorithmicity of the mind, forms the foundation for Penrose's whole argument. Accordingly, he devotes over half of his first book to proving it. It goes something like this:

1a. Assume that my powers of mathematical reasoning can be captured by some formal system F—say the algorithms of a computer.

1b. Within this formal system F I can find a knowably sound algorithm for doing mathematics. I'll call this algorithm the "Gödel sentence" G(F). To give an idea of what this Gödel sentence looks like, let me just say for now that it resembles the following sentence: *This sentence cannot be proven.*

1c. Now, I know the Gödel sentence G(F) to be true. Still, I cannot show this fact within my formal system, due to G(F)'s peculiar recursive nature.

1d. Furthermore, if I expand my formal system F to include G(F) as an axiom, this larger formal system—let's call it F*—has its own true but undecidable Gödel sentence G(F*).

1e. Therefore, my powers of mathematical reasoning cannot be captured by F, F*, or any arbitrarily large formal system of algorithms.

1f. Therefore, I am not a computer.

Now, obviously Penrose's proof is far more detailed than this. If you have any interest in it, you can pick up from me at the end of this talk a short handout that goes through the details. I used the simplest version I could find that still preserves Penrose's rigor.[9] But I think even this brief exposition conveys the gist of this part of his argument.

Let's now consider Penrose's second point, namely, that consciousness is a reality that deserves scientific explanation. Here Penrose does not mention much more than his unwillingness to consider the alternative. It may well be, as du Bois-Reymond suggested, that no theory of science will ever explain the mind, but this is not a possibility that Penrose wishes to entertain. To him it smacks of mysticism. Nevertheless, Penrose finds contemporary physical theory unsatisfactory. Several deficiencies spring to his mind. First, the two greatest achievements in twentieth century physics, general relativity and quantum mechanics, are incommensurable. Each has withstood all challenge individually, but they cannot both be true. Second, no physical theory, not even the standard explanations given in statistical mechanics, can explain the second law of thermodynamics, which states that the entropy of the cosmos is increasing. In other

[9] To me the best summaries and critiques are Daryl McCullough, "Can Humans Escape Gödel? A Review of Shadows of the Mind by Roger Penrose," *Psyche* 2.4 (April 1995); David J. Chalmers, "Minds, Machines, and Mathematics: A Review of *Shadows of the Mind* by Roger Penrose," *Psyche* 2.9 (June 1995); Rick Grush and Patricia Smith Churchland, "Gaps in Penrose's Toilings," *Journal of Consciousness Studies* 2.1(1995): 10-29; Geoffrey LaForte, Patrick J. Hayes, Kenneth M. Ford, "Why Gödel's Theorem Cannot Refute Computationalism," *Artificial Intelligence* 104 (1998): 265-286.

words, we do not understand why time runs in one direction. Finally, no full account exists for the reduction of the state vector, or collapse of the wave function, in the measurement of quantum events. Several interpretations have been proposed, but each leads to what Penrose terms "puzzles" and "paradoxes." After all, a fully scientific model of nature would have to incorporate the fact of consciousness.[10]

Penrose develops this last point at briefly in his first book and at greater length in *Shadows of the Mind*. This is where he becomes most speculative, replacing one mystery—consciousness—with two: quantum gravity and neural microtubules. His reasoning goes a little like this:

4a. The mind must have a means of interacting with matter.

4b. This process must be both nonalgorithmic (to avoid Gödel's pitfalls) and determinist (to avoid reducing thought to magic).

4c. A correct theory of quantum gravity could provide a determinist, nonalgorithmic account of this mind/body interaction.

4d. Such an interaction would have to take place in the brain in structures small enough to preserve quantum effects.

4e. Microtubles within neurons serve as a possible site for this to occur.

Criticism of Penrose's Thesis

As you may well imagine Penrose has received his fair share of criticism. Most objections to his thesis fall into four camps. These could be classified under the rubrics

[10] More discussion of these issues can be found in Euan J. Squires, *Conscious Mind in the Physical World* (Bristol; New York: Hilger, 1990); Henry P. Stapp, *Mind, Matter, and Quantum Mechanics* (Berlin: Springer-Verlag, 1993).

Logical, *Physical*, *Biological*, and *Philosophical*. I wish to suggest a fifth rubric of my own: *Historical*. But first, let me try to work through the contemporary polemic.

Logical. This might be summed up as "Penrose has misunderstood Gödel." Some of the arguments made here go right over my head. But basically, they all seem to come back to one central point: Gödel proved that no *consistent* formalism is *complete*. So either Penrose's argument is incomplete, in which case it is false, or it is inconsistent, in which case it is equivocal. I will return to this point in a minute.[11]

Physical. This might be summed up as "Penrose has misunderstood quantum mechanics." Most criticism focuses on Penrose's view that a correct theory of quantum gravity could account for an "objective reduction" of the wave function. Present theories ascribe this decoherence to "interactions with the environment or … fluctuations in the topology of spacetime."[12] But, according to Stephen Hawking,

> Roger seems to want neither of these mechanisms. Instead he claims that it occurs because of the slight warping of space-time produced by the mass of a small object. … That warping will not prevent a Hamiltonian evolution with no decoherence or objective reduction. It may be that accepted ideas are wrong but Roger has not put forward a detailed theory that would enable us to calculate when objective reduction would occur.[13]

Biological. This might be summed up as "Penrose has misunderstood evolution and neuroscience." Typical critiques point out that natural selection rewards heuristics, not proofs, that much of the brain works unconsciously, preconsciously, or emotionally, and that quantum processes do not seem to function in organisms.

Philosophical. This might be summed up as "Penrose has misunderstood consciousness and reductionism." Many philosophers believe that the privacy of consciousness

---

[11] My favorite Gödelian objection goes like this: Penrose's argument for the nonalgorithmicity of thought is entirely formal. One could imagine it a planet of intelligent robots making it.
[12] Stephen Hawking, "The Objections of an Unashamed Reductionist," in *The Large, The Small, and the Human Mind*, op. cit., 169-172, on 170.
[13] Ibid.

precludes it from scientific investigation. If we had machines that allowed us to read each other's thoughts, we might be able to experiment on them. But until that time, cognitive research remains confined to public phenomena like intelligence. Moreover, there is no reason to assume that consciousness will reduce to physics. It may be an entity with its own intrinsic qualities, or it may make most sense within the disciplines of psychology and physiology. Penrose assumes no explanation to be fully competent until it is freed of context and expressed in mathematics.[14]

Conclusion: Platonism, or the Religion of Science

Universal, eternal truth has often been equated with God. Penrose equates it with mathematics. In fact, he is so certain that mathematics reveals a perfect world that he wrote three books justifying this belief.

The great irony is that Penrose adapted Gödel's proof to his endeavor. Most logicians interpret Gödel's result as a statement on limitation, one of the latest in a long history of failures to find a firm foundation to mathematics. That history leads back to the set theory of Whitehead and Russell and Cantor, to the geometry of Riemann and Bolyai and Lobachevsky, to the analysis of Weierstrass and Cauchy and Lagrange, all the way to the Ancient Greeks. Limitation is old. It is also common. In fact, it could be considered the major theme of the twentieth century: in politics, with failures of Hitler and Stalin and Mao; in social science, with the rise of complexity and context; in philosophy, with the silence of Wittgenstein; in literature, with the novels of Joyce and Proust and Musil; in art, with the end of style; in music, with the elision of composition and performance; in

[14] Nancy Cartwright, "Why Physics?" in *The Large, The Small, and the Human Mind*, op. cit., 161-168.

science and technology, with the loss of public trust; and in medicine; with the recrudescence of disease.

Some time ago Owen Chadwick made the brilliant observation that the first wave of secularization in the nineteenth century mirrored the very religion it was attempting to replace. Du Bois-Reymond belonged to that first wave. He was born in 1818, the same year as Karl Marx, Ivan Turgenev, James Froude, and Frederick Douglass. These contemporaries received educations that Flaubert termed sentimental. I suppose that is a generous way to speak of lives shaped by disappointment. The thing is, limitation also teaches acceptance. This is not a lesson Roger Penrose seems ready to learn.